Language Learning in Multi-Agent Systems

Martin Allen
Computer Science Department
University of Massachusetts
Amherst, MA 01003, USA
mwallen@cs.umass.edu

Claudia V. Goldman
Caesarea Rothschild Institute
University of Haifa
Mount Carmel, Haifa 31905, Israel
clag@cri.haifa.ac.il

Shlomo Zilberstein
Computer Science Department
University of Massachusetts
Amherst, MA 01003, USA
shlomo@cs.umass.edu

Abstract

We present the problem of learning to communicate in decentralized and stochastic environments, analyzing it formally in a decision-theoretic context and illustrating the concept experimentally. Our approach allows agents to converge upon coordinated communication and action over time.

1 Introduction

Learning to communicate in multi-agent systems is an emerging challenge AI research. Autonomous systems, developed separately, interact more and more often in contexts like distributed computing, information gathering over the internet, and wide-spread networks of machines using distinct protocols. As a result, we foresee the need for autonomous systems that can learn to communicate with one another in order to achieve cooperative goals. We make some first steps towards solving the attendant problems.

Coordination among agents acting in the same environment while sharing resources has been studied extensively, particularly by the multi-agent systems community. While such coordination may involve communication, typically there is no deliberation about the value of communication, resulting in systems with no communication or ones allowing free communication of well-understood messages. In contrast, we study decentralized systems that require agents to adapt their communication language when new situations arise or when mis-coordination occurs.

2 The Decentralized Learning Framework

We study the problem in the context of *decentralized Markov Decision Processes* [Bernstein *et al.*, 2002] with communication (Dec-MDP-Com). Such a process is a multi-agent extension of a common MDP in which each agent α_i observes only its own local portion of the state-space, and can attempt to communicate with others using the set of messages Σ_i. Decentralization makes Dec-MDPs, with communication or not, significantly harder to solve than regular MDPs; for their complexity properties, see [Goldman and Zilberstein, 2004].

If agents in a system share the same language, optimal linguistic action is a matter of deciding *what* and *when* to communicate, given its cost relative to the projected benefit of sharing information. However, where agents utilize different sets of messages, and do not fully understand one another, message-passing alone is not enough. Rather, agents need to learn how to *respond* to the messages that are passed between them—in a sense, learning what those messages *mean*.

Definition 1 (Translation). Let Σ and Σ' be sets of messages. A *translation*, τ, between Σ and Σ' is a probability function over message-pairs: for any messages σ, σ', $\tau(\sigma, \sigma')$ is the probability that σ and σ' have the same meaning. $\tau^{+}_{\Sigma,\Sigma'}$ is the set of all translations between Σ and Σ'.

Agents may need to consider multiple possible translations between messages; that is, agents possess beliefs as to which translation is correct given their present situation.

Definition 2 (Belief-state). Let agents α_1, α_2 use sets of messages Σ_1, Σ_2. A *belief-state* for α_i is a probability-function β_i over translation-set $\tau^{+}_{\Sigma_i,\Sigma_j}$ $(i \neq j)$. For translation $\tau \in \tau^{+}_{\Sigma_i,\Sigma_j}$, $\beta_i(\tau)$ is the probability that τ is correct.

Updating beliefs about translations is thus an important part of the overall process of learning to communicate. Agents act based upon local observations, messages received, and current beliefs about how to translate those messages. Their actions lead to new observations, causing them to update beliefs and translations. The procedure governing these updates comprises the agent's *language-model*, a function from actions, messages, and observations to distributions over translations. Such models may be highly complex, or difficult to compute, especially where languages are complicated, or the environment is only partially observable. Here we concentrate upon special—but interesting—cases for which generating these probabilities is much more straightforward.

3 Formal Properties of the Problem

Our main formal results isolate conditions under which Dec-MDP-Coms reduce to simpler problems, and present a protocol for learning to communicate in such reduced problems.

Reduction to MMDPs

Boutilier [1999] defines *multiagent MDPs* (MMDPs), consisting of a set of agents operating in a fully- and commonly-observed environment; transitions between states in that environment arise from *joint actions* of all agents, and a common reward is shared by the system as a whole. While we can

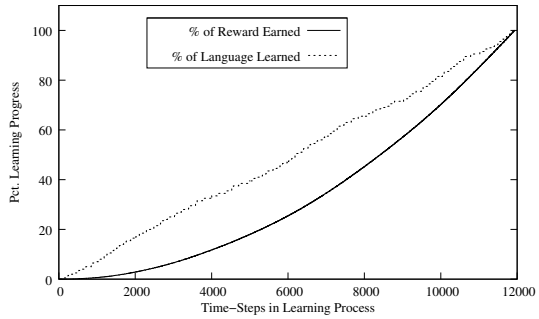

Figure 1: Reward accumulated as language is learned.

calculate an optimal joint policy for such a process offline, this is not the same thing as *implementing* it. Unless agents can coordinate their actions, there is no guarantee of a jointly optimal policy, since communication is not allowed, or is unreliable. Boutilier thus defines *coordination problems*, which arise when agents may each take an individual action that is potentially optimal, but which combine in sub-optimal fashion. We show that certain, putatively more complex, Dec-MDP-Coms in fact reduce to MMDPs for which such problems do not arise. This is notable, as Dec-MDPs are generally intractable, while MMDPs can be solved efficiently.

Definition 3 (Fully-describable). A Dec-MDP-Com is *fully-describable* if and only if each agent α_i possesses a language Σ_i that is sufficient to communicate both: (*a*) any observation it makes, and (*b*) any action it takes.

Definition 4 (Freely-describable). A Dec-MDP-Com is *freely-describable* if and only if for any agent α_i and message $\sigma \in \Sigma_i$, the cost of communicating that message is 0.

Claim 1. A Dec-MDP-Com is equivalent to an MMDP without coordination problems if (*a*) it is both fully- and freely-describable; and (*b*) agents share a common language. □

Suitability and Convergence

For any freely- and fully-describable Dec-MDP-Com, agents can calculate an optimal joint policy, under the working assumption that all agents share a common language and that all relevant information is shared. Where agents must in fact learn to communicate, however, implementation of such policies requires cooperation from the environment, so that agents can update translations appropriately over time. The full definition of a *suitable* Dec-MDP-Com cannot be included here; we simply note that in such problems, the probability that each agent assigns to the actual prior observations and actions of others *following* some state-transition is strictly greater than that of the observations and actions considered most likely *before* that transition (unless those entries were actually correct). Suitable Dec-MDP-Coms provide enough information to ensure that others' actual actions and observations are more likely than mistaken ones.

We extend work of Goldman *et al*. [2004] (where agents communicate states but not actions), to give an *elementary action protocol*. Using such a protocol for action and belief-update, agents move towards optimality, based upon the observed consequences of action in a suitable problem-domain.

Claim 2. Given an infinite time-horizon, agents acting according to the elementary action protocol in a suitable Dec-MDP-Com will eventually converge upon a joint policy that is optimal for the states they encounter from then on. □

4 Empirical Results and Conclusions

To explore the viability of our approach, we implemented our language-learning protocol for a reasonably complex (but still suitable) Dec-MDP-Com, involving two agents in joint control of a set of pumps and flow-valves in a factory setting.

Our results show the elementary protocol converging on optimal policies in a wide range of problem-instances. Figure 1 gives an example, for a problem-instance featuring 100 vocabulary-items for each agent, showing the percentage of total accumulated reward, and total shared vocabulary, at each time-step in the process of learning and acting. As can be seen, the learning process (top, dotted line) proceeds quite steadily. Reward-accumulation, on the other hand, grows with time before finally stabilizing. Initially, language learning outpaces reward gain given that knowledge, as agents still find many of the other's actions and observations hard to determine. As time goes on, the rate of accumulated reward narrows this gap considerably; agents now know much of what they need to communicate, and spend more time accumulating reward in familiar circumstances, without necessarily learning anything new about the language of others.

These experimental results conform with intuition, showing that while a small amount of language learning does little to help agents in choosing their actions, they are capable of very nearly optimal action even in the presence of an understanding that is still less than perfect. This opens the door for further study into approximation in these contexts. We continue to investigate and compare other approaches to the problem, including analysis of the differences between possible optimal offline techniques and online learning methods.

Acknowledgments

This work was supported in part by the National Science Foundation under grant IIS-0219606 and by the Air Force Office of Scientific Research under grant F49620-03-1-0090.

References

[Bernstein *et al.*, 2002] D. S. Bernstein, R. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of Markov decision processes. *Mathematics of Operations Research*, 27:819–840, 2002.

[Boutilier, 1999] C. Boutilier. Sequential optimality and coordination in multiagent systems. In *Procs. of IJCAI-99*, pages 478–485, Stockholm, Sweden, 1999.

[Goldman and Zilberstein, 2004] C. V. Goldman and S. Zilberstein. Decentralized control of cooperative systems: Categorization and complexity analysis. *Journal of Artificial Intelligence Research*, 22:143–174, 2004.

[Goldman *et al.*, 2004] C. V. Goldman, M. Allen, and S. Zilberstein. Decentralized language learning through acting. In *Procs. of AAMAS-04*, pages 1006–1013, New York City, NY, 2004.